Definitions: Let A be the set of all actions that can be sampled with PASPO, and let $P = \{a \in \mathbb{R}^n | Ca \leq b\}$ be the convex polytope that corresponds to constrained action space. We define $A_*^{(i+1)} = \{a \in \mathbb{R}^n | C^{(i+1)} a^{(i+1)} \leq b_*^{(i+1)}, \forall j = 1, \dots, i : a_j = a_j^*\}$ where $b_*^{(i+1)} = b - \sum_{j=1}^{i} a_j^* \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix}$ and $C^{(i+1)}$ and $a^{(i+1)}$ as defined in the paper.

Thus, $A_*^{(i+1)}$ is the restricted action space after sampling/fixing already the allocations $a_1^*, \dots, a_i^*$.

Theorem 1. *Let $P = \{a \in \mathbb{R}^n | Ca \leq b\} \neq \emptyset$ be the convex polytope that corresponds to a constrained action space. Let A be the set of all the points that can be generated by PASPO. It holds that $A = P$.*

Proof. Well-defined: Show that $A^{(n)} \neq \emptyset$ if $P \neq \emptyset$.

Induction over i:

$i = 1:$ $\quad A_*^{(1)} = \{a \in \mathbb{R}^n | C^{(1)} a^{(1)} \leq b_*^{(i+1)}\} = \{a \in \mathbb{R}^n | Ca \leq b\} = P \neq \emptyset$

$i \to i+1:$

$(i + 1 \leq n)$ $\quad A_*^{(i)} \neq \emptyset \Rightarrow \exists a^{\uparrow}, a^{\downarrow} \in A_*^{(i)} : a_i^{\uparrow} = a_i^{\min}, a_i^{\downarrow} = a_i^{\max}$

Now assume an arbitrary a_i^* is sampled from $[a_i^{\min}, a_i^{\max}]$

$\Rightarrow \exists \lambda \in [0, 1] : a_i^* = (\underbrace{\lambda a^{\downarrow} + (1 - \lambda) a^{\uparrow}}_{:= a^{\lambda}})_i$

By convexity of polytopes as solution spaces for linear inequality systems, we get:

$$\begin{bmatrix} c_{1,i} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{m,i} & \cdots & c_{m,n} \end{bmatrix} \begin{bmatrix} a_i^{\lambda} \\ \vdots \\ a_n^{\lambda} \end{bmatrix} \leq b - \sum_{j=1}^{i-1} a_j^* \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix} \underset{(a_i^{\lambda} = a_i^*)}{\Longleftrightarrow}$$

$$\begin{bmatrix} c_{1,i+1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{m,i+1} & \cdots & c_{m,n} \end{bmatrix} \begin{bmatrix} a_{i+1}^{\lambda} \\ \vdots \\ a_n^{\lambda} \end{bmatrix} \leq b - \sum_{j=1}^{i} a_j^* \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix} \Rightarrow a^{\lambda} \in A_*^{(i+1)}$$

To show that $A = P$:

$A \subseteq P:$ $\quad$ Let $a^* \in A$. In the last step (n), a_n^* is sampled (by design) such that

$$C^{(n)} a_n^* \leq b - \sum_{j=1}^{n-1} a_j^* \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix} \Leftrightarrow Ca^* \leq b \Leftrightarrow a^* \in P$$

$A \supseteq P:$ $\quad$ Let $a^* \in P$. $\Leftrightarrow Ca^* \leq b \Leftrightarrow C^{(i)} a^* \leq b^{(i)}\ \forall i \Leftrightarrow a^* \in A_*^{(i)}\ \forall i$

$\Rightarrow$ We can construct a^* by sampling a_i^* in every step i. $\Rightarrow a^* \in A$

$\square$

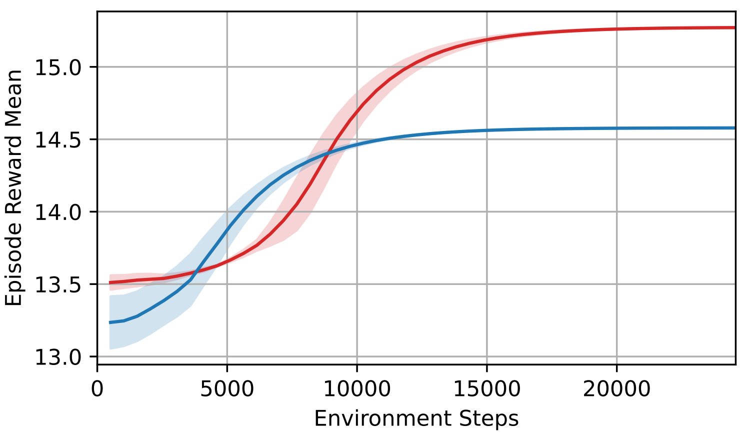

Figure 1: The impact of the allocation order on PASPO without de-biased initialization in the synthetic benchmark with two states, a 7-dimensional action space, and no additional allocation constraints. Blue depicts the standard allocation order (i.e., $e_1, e_2, \dots, e_n$) and red depicts the reversed allocation order (i.e., the entities are allocated in the reversed order). A significant difference in performance can be observed with respect to the order without our de-biased initialization. In contrast, Figure 5b in the paper shows that with the de-biased initialization the difference is not significant.